

The thesis deals with language modelling for German. The main concerns are the specifics of German language that are troublesome for standard n-gram models. First the statistical methods of language modelling are described and language phenomena of German are explained. Following that suggests own variants of n-gram language models with an aim to improve these problems. The models themselves are trained using the standard n-gram methods as well as using the method of maximum entropy with n-gram features. Both possibilities are compared using correlation metrics of hand-evaluated fluency of sentences and automatic evaluation - the perplexity. Also, the computation requirements are compared. Next, the thesis presents a set of own features that represent the count of grammatical errors of chosen phenomena. Success rate is verified on ability to predict the hand-evaluated fluency. Models of maximum entropy and own models that classify only using the medians of phenomena values computed from training data are used.